

Bausteine Forschungsdatenmanagement
Empfehlungen und Erfahrungsberichte für die Praxis von
Forschungsdatenmanagerinnen und -managern

Datenmanagement im SFB 1313

Martin Schneider Bernd Flemisch Steffen Frey
Sibylle Hermann Dorothea Iglezakis Matthias Ruf
Björn Schembera Anett Seeland Holger Steebⁱ

2020

Zitiervorschlag

Schneider, Martin et al.. 2020. Datenmanagement im SFB 1313. *Bausteine Forschungsdatenmanagement. Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanagerinnen und -managern* Nr. 1/2020: S. 28-38. DOI: [10.17192/bfdm.2020.1.8085](https://doi.org/10.17192/bfdm.2020.1.8085).

Dieser Beitrag steht unter einer
[Creative Commons Namensnennung 4.0 International Lizenz \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/).

ⁱAlle Autorinnen und Autoren gehören verschiedenen Instituten und Einrichtungen der Universität Stuttgart an. ORCID: Martin Schneider ([0000-0002-4128-4930](https://orcid.org/0000-0002-4128-4930)), Bernd Flemisch ([0000-0001-8188-620X](https://orcid.org/0000-0001-8188-620X)), Steffen Frey ([0000-0002-1872-6905](https://orcid.org/0000-0002-1872-6905)), Sibylle Hermann ([0000-0001-9239-8789](https://orcid.org/0000-0001-9239-8789)), Dorothea Iglezakis ([0000-0002-8524-0569](https://orcid.org/0000-0002-8524-0569)), Matthias Ruf ([0000-0003-0299-5921](https://orcid.org/0000-0003-0299-5921)), Björn Schembera ([0000-0003-2860-6621](https://orcid.org/0000-0003-2860-6621)), Anett Seeland, Holger Steeb ([0000-0001-7602-4920](https://orcid.org/0000-0001-7602-4920)).

1 Abstract

Dieser Artikel gibt einen Überblick über das Forschungsdatenmanagement im SFB 1313. Ein wesentliches Merkmal der geplanten Forschungstätigkeit im SFB ist die Verknüpfung von physikalischen und mathematischen Modellen und den daraus resultierenden Rechenmodellen mit hochaufgelösten Experimenten. Eine solche Verknüpfung stellt diverse Anforderungen an das Forschungsdatenmanagement. Diese Anforderungen sowie die damit einhergehenden Herausforderungen werden in diesem Artikel detailliert beschrieben. In diesem Zusammenhang wird das Datenrepositorium DaRUS vorgestellt, welches die Verwaltung und Beschreibung der im SFB anfallenden Daten ermöglicht. Zusätzlich wird das entwickelte Metadatenschema vorgestellt und auf die geplante automatisierte Metadatenerfassung eingegangen.

2 Einführung

In diesem Artikel soll ein Überblick über die geplanten und bereits realisierten FDM-Maßnahmen im Sonderforschungsbereich (SFB) 1313¹ gegeben werden. Der SFB 1313 trägt den Titel "Grenzflächengetriebene Mehrfeldprozesse in porösen Medien - Strömung, Transport und Deformation". Poröse-Medien-Systeme haben in vielfältigen Anwendungsbereichen, unter anderem in ökologischen, technischen und biologischen Systemen, eine hohe Relevanz. Klimamodelle, Brennstoffzellen, Kühlsysteme, Energiespeicherung im natürlichen Untergrund, Tumorwachstum, Medikamententransport in menschlichen Zellen oder Vertebroplastie sind nur einige Beispiele. Insbesondere haben Grenzflächen einen großen Einfluss auf Mehrfeldprozesse wie Strömung, Transport und Deformation in solchen Systemen. Der SFB 1313 hat zum Ziel, diese Grenzflächen zu erforschen und ein grundlegendes Verständnis darüber zu gewinnen, wie sie sich auf Mehrfeldprozesse auswirken. Er ist ein interdisziplinärer Sonderforschungsbereich der Universität Stuttgart, der aus vier großen Projektbereichen (A–D) und 17 untergeordneten Einzelprojekten besteht: A "Freie Strömung und Poröse-Medien-Strömung," B "Rissausbreitung und Fluidströmung," C "Fluid-Festphasen /änderung" sowie D "Benchmarks, Computing und Visualisierung". Der SFB 1313 startete Anfang 2018, wird von der Deutschen Forschungsgemeinschaft (DFG) gefördert und ist dem Stuttgarter Zentrum für Simulationswissenschaft (SC SimTech)² angegliedert. Rund 50 Forscherinnen und Forscher aus 18 verschiedenen Instituten der Universität Stuttgart und Partneruniversitäten bilden das Team des SFB 1313. Ihre klassischen Forschungsfelder reichen von Ingenieursdisziplinen wie Thermodynamik, Festkörper- und Fluidmechanik über Informatik bis hin zu Mathematik und Physik.

Ein wesentliches Merkmal der geplanten Forschungstätigkeit im SFB 1313 ist die Integration von physikalischen und mathematischen Modellen und den daraus resultie-

¹<https://www.sfb1313.uni-stuttgart.de/>. Letzter Zugriff 30.04.2019

²https://www.simtech.uni-stuttgart.de/ueber_uns/SC-SimTech/. Letzter Zugriff 30.04.2019

renden Rechenmodellen mit hochauflösten Experimenten. Wie in [Abbildung 1](#) exemplarisch dargestellt, gehen damit vielfältige Anforderungen an das entsprechende Forschungsdatenmanagement einher, insbesondere die Verknüpfung experimenteller und simulativer Daten. Um diesen Anforderungen gerecht zu werden, wurde von der Universität Stuttgart die Stelle eines/r Daten- und Softwaremanagers/in geschaffen, die über die gesamte Förderdauer des SFBs hinweg finanziert wird. Die Stelle ist am Institut des Sprechers angesiedelt und dafür verantwortlich, Konzepte für FDM und Softwareentwicklung innerhalb des SFBs zu entwickeln und zusammen mit den Forschenden umzusetzen.

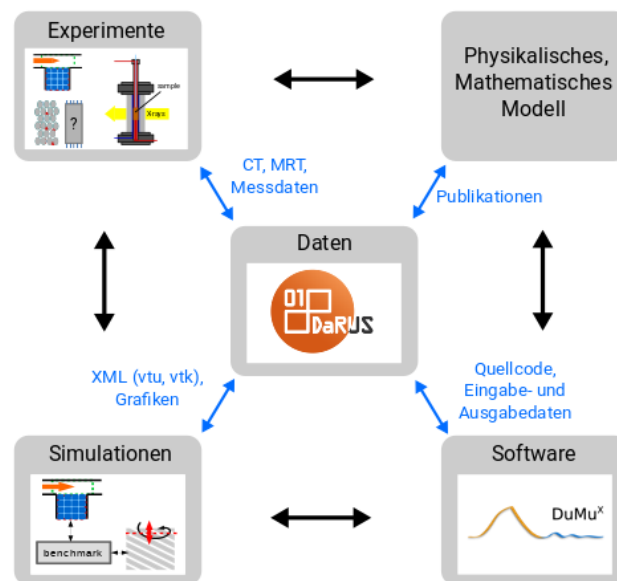


Abbildung 1: Forschungsdaten-Workflow im SFB 1313

Zur Illustration werden zwei klassische Anwendungsfälle innerhalb des SFBs vorgestellt, die beide versuchen, Experiment mit Simulation zu verbinden. Innerhalb des ersten Use-Cases liegt der Fokus auf der Validierung von Simulation durch Experiment. Der experimentellen Datengewinnung liegen hier 2D-Mikrofluidikexperimente zugrunde. Kennzeichnend hierfür ist die Datenerfassung im Zeitbereich. Neben der reinen Validierung sind experimentell gewonnene Bilddaten oftmals auch die direkten Eingabedaten von Simulationen. Klassisch hierfür ist die hochauflösende räumliche Erfassung von inneren Strukturen (morphologische Daten des Porenraums oder der Verteilung der Fluidphasen bei Mehrphasenstroemungen) durch μ CT-Scans, was den zweiten Use-Case darstellt.

Aufgrund der in beiden Fällen sehr großen Datenmengen von teils über 100 Gigabyte pro Datei, stellen der Datenaustausch und die Datenarchivierung besondere Herausforderungen dar. Desto mehr rückt die Erfassung von Metadaten in den Mittelpunkt,

die die Reproduzierbarkeit und Vergleichbarkeit von Experiment und Simulation sowie die spätere Suche nach Datensätzen durch Filterwerkzeuge überhaupt erst ermöglicht.

Der Austausch, die Beschreibung und die Verwaltung von Datensätzen soll über das Datenrepositorium der Universität Stuttgart (DaRUS) als zentrales Bindeglied erfolgen, worüber auch Datensätze direkt publiziert werden können. Als wesentlich wird hierbei die Automatisierung der Datenprozesse angesehen. Die Automatisierung umfasst dabei zunächst die Erfassung aller relevanter Eingabe- und Ausgabedaten, die das jeweilige Experiment oder die Simulation beschreiben. In einem weiteren Schritt werden diese Daten automatisiert mit dem Metadatenschema EngMeta beschrieben (siehe Abschnitt 4.2 und gemeinsam mit den Metadaten auf DaRUS abgelegt. Für den automatisierten Upload werden APIs des zugrundeliegenden Systems Dataverse verwendet. Die hochgeladenen Datensätze sollen zusätzlich mit Visualisierungen angereichert werden, die eine schnelle Voransicht der Daten erlauben.

3 Datenarten/-menge und Herausforderungen

Die im SFB anfallenden Daten sind exemplarisch in [Abbildung 1](#) dargestellt und werden im Folgenden beschrieben. Hierbei sind die größten Herausforderungen zum einen die automatisierte Metadatenerfassung und zum anderen der Upload von großen Datenmengen. Diese Herausforderungen werden in diesem Abschnitt nur kurz skizziert, für weitere Details sei auf die Abschnitte 4.3 und 4.1 verwiesen. [Tabelle 1](#) gibt außerdem einen Überblick über die verschiedenen Datentypen und die im SFB anfallenden Datenmengen.

3.1 Experimentelle Daten

Mikrofluidikexperimente erlauben eine bildgebende Charakterisierung der Strömungsprozesse in komplexen zweidimensionalen Porennetzwerken. Der Ausgangspunkt von Mikrofluidikuntersuchungen ist ein CAD-basiertes Design des Porenraums eines charakteristischen porösen Mediums. Durch die Kombination von foto- und softlithografischen Verfahren kann damit nach verschiedenen Herstellungsschritten eine Mikrofluidikzelle aus Polydimethylsiloxan (PDMS) hergestellt werden³. Die PDMS-Zelle ist optisch transparent und kann deshalb optimal für bildgebende strömungsmechanische Untersuchungen (optische Mikroskopie, Particle Imaging Velocimetry (μ PIV)) verwendet werden. Die optischen Mikroskopieverfahren erzeugen dabei Datensätze aus Bildern von jeweils ca. 5 MPixel mit einer Bildrate von bis zu 1000 fps. Typische Untersuchungen sind sogenannte Drainage-Imbibition-Cycles. Um solche Prozesse charakte-

³N. K. Karadimitriou und S. M. Hassanizadeh. 2012. "A Review of Micromodels and Their Use in Two-Phase Flow Studies." *Vadose Zone Journal* 11(3). <https://doi.org/10.2136/vzj2011.0072>.

risieren zu können, muss eine größere Anzahl an Zyklen durchlaufen werden, wodurch die Untersuchungszeit mehrere Stunden betragen kann.

Durch die Computertomographie (XRCT) ist es möglich, die innere Struktur von Objekten zu untersuchen und ein digitales Abbild dieser Struktur zu erstellen. Nach entsprechender Datenaufbereitung kann dieses Abbild als nahezu reale Ausgangsgeometrie für Simulationen genutzt werden. Hierfür werden in der Computertomographie sogenannte Absorptionsprofile des Untersuchungsobjekts aus vielen Richtungen erstellt, um daraus im Anschluss die Volumenstruktur zu rekonstruieren. Hinsichtlich der entstehenden Daten muss zwischen Roh- und rekonstruierten Datensätzen differenziert werden. Die Größe der Rohdaten ist abhängig von der Auflösung und Pixeltiefe des genutzten Flachbilddetektors und der Anzahl der durchgeführten Projektionen. Typische Rohdatensätze liegen im Bereich von wenigen Gigabyte bis in den teils hohen zwei- und sogar dreistelligen Gigabyte-Bereich. Mittels der gefilterten Rückprojektion wird im Anschluss aus den Rohdaten über verschiedene Zwischenschritte die 3D-Struktur rekonstruiert, wobei die Ergebnislösung und damit das Entstehen von Artefakten stark mit der Anzahl an Projektionen gekoppelt ist. Die Größe rekonstruierter Datensätze, die aus einem Stack von Grauwert-Schnittbildern bestehen, hängt wiederum von der Auflösung des eingesetzten Flachbilddetektors und dessen Pixeltiefe ab und liegt hier im Bereich von 10 GB bis 40 GB.

Die Transparenz und die Reproduzierbarkeit der Datensätze kann nur gewährleistet werden, wenn neben den reinen Bilddatensätzen weitere Informationen vorliegen. Notwendige Informationen resultieren unter anderem aus eingesetzter Hard- und Software mitsamt den verwendeten Parametern, beispielsweise des Flachbilddetektors und der Röntgenquelle oder die Probenpositionierung. Diese Informationen sollen für das jeweilige Experiment weitestgehend automatisiert erfasst werden und bei der Archivierung des Datensatzes beziehungsweise der Verlinkung in DaRUS entsprechend dem Metadatenschema in [Abbildung 2](#) abgelegt werden.

3.2 Simulationsdaten

Die Simulation verschiedener Anwendungen basiert auf physikalischen und mathematischen Modellen und den daraus resultierenden Rechenmodellen. Hierfür werden im SFB 1313 ausschließlich gitterbasierte Methoden (Diskretisierungsverfahren) verwendet. Basierend auf hochauflösenden μ CT-Scans zur räumlichen Erfassung von Strukturen (siehe letzter Abschnitt) wird ein Gitter erzeugt, welches die darstellende Geometrie approximiert. Je nach Ansatz müssen Poren detailliert aufgelöst oder durch einen Mittelungsansatz beschrieben werden. Hierfür müssen physikalisch relevante Modellparameter basierend auf experimentellen Daten bestimmt werden. Experimente liefern daher essentielle Eingabedaten für die Rechenmodelle. Sämtliche Eingabedaten, wie beispielsweise das Rechengitter, und Modellparameter sind Metadaten, die entsprechend automatisiert erfasst werden sollen (siehe Abschnitt 4.3). Die Erfassung

aller modellrelevanter Parameter, welche während des Simulationslaufs verwendet werden, erlaubt die Reproduzierbarkeit von Simulationsdaten (siehe Abschnitt 3.3).

Die Ausgabe von Simulationsergebnissen erfolgt hauptsächlich im XML-Format (vtu und vtk Dateien). Hierbei werden physikalische Größen zu unterschiedlichen Zeitpunkten mit dem zugrundeliegenden Rechengitter verknüpft. Die Nachbearbeitung und Visualisierung erfolgt dann entweder mit Softwarelösungen, wie beispielsweise Paraview, oder direkt auf DaRUS (siehe Abschnitt 4.4). Durch Mikrofluidikexperimente lassen sich die Simulationsergebnisse dann entsprechend validieren.

Es ist auch geplant, dass experimentelle Daten direkt während des Simulationslaufs integriert werden, um beispielsweise Modellparameter anzupassen. Wie bereits erwähnt ist hierbei die automatisierte Metadatenerfassung entscheidend.

3.3 Softwarestacks

Um die Reproduzierbarkeit von Simulationsdaten zu gewährleisten, sollen sogenannte Softwarestacks gespeichert werden, welche die kompletten ausführbaren Umgebungen, die zur Durchführung der Simulationen notwendig sind, in Form von beispielsweise Docker-Containern bereitstellen. Zusätzlich wird der Quellcode der verwendeten Softwarelösungen mittels GitLab archiviert und entsprechend auf DaRUS verlinkt. Hierfür werden entsprechende Metadatenfelder in DaRUS bereitgestellt, siehe [Abbildung 2](#). Dadurch lassen sich der Quellcode und die Softwarestacks direkt mit den Simulationsergebnissen verknüpfen.

	Datentyp	Dateigröße	Gesamt
Experiment	Bild- und Videoformate	<150 GB	100 TB
Simulation	XML, VTK, Bildformate	<10 GB	40 TB
Softwarestacks	Container	<2 GB	<1 TB
Visualisierung	Bild- und Videoformate	100 MB -1 GB	10 TB

Tabelle 1: Datentypen

4 Geplante Umsetzung

Die oben beschriebenen Daten sollen in einem System verwaltet werden, das eine differenzierte Suche nach den Daten und ihren Eigenschaften ermöglicht, die verschiedenen Datensätze miteinander verknüpft, die Datensätze durch eine detaillierte Beschreibung des Entstehungsprozesses verständlich und nachvollziehbar macht und den Umgang mit großen Datenmengen ermöglicht. Gleichzeitig soll die Eingabe, Beschreibung und Visualisierung der Daten so weit wie möglich automatisiert werden und dennoch in einer strukturierten und standardisierten Weise geschehen.

Diese Anforderungen sollen durch verschiedene Komponenten umgesetzt werden - das Datenrepositorium DaRUS zur durchsuchbaren Verwaltung der Daten, das Metadatenschema EngMeta zur strukturierten Beschreibung und eine Automatisierung der Erfassung bereits vorhandener Metadateninformationen, sowie der Visualisierung der Daten.

4.1 DaRUS - das Datenrepositorium der Universität Stuttgart

DaRUS, das Datenrepositorium der Universität Stuttgart, dient der Verwaltung und Beschreibung von Daten und Modellen im Datenlebenszyklus. DaRUS basiert auf Dataverse⁴, einer OpenSource-Software entwickelt vom IQSS Harvard (gemeinsam mit einer großen und aktiven internationalen Entwicklergemeinschaft). Dataverse bietet ein feingranulares Rechte- und Rollen-Management, das auf Bereiche - sogenannte Dataverses - angewendet wird. Damit ist es möglich Daten selbstbestimmt zu teilen und zu verwalten, ohne diese gleich veröffentlichen zu müssen. So können die Forschungsdaten aus dem SFB 1313 bereits im Forschungsprozess mit auf den Fachbereich angepassten Metadaten versehen werden. Während Dataverse eine durchsuchbare Datenbank für die Metadaten darstellt, liegen die eigentlichen Dateien in einem Object-Store, auf den Dataverse über eine S3-Schnittstelle zugreift.

Dataverse adressierte ursprünglich vornehmlich die Sozialwissenschaften. Deshalb ist die Unterstützung großer Datenmengen noch rudimentär. Aus den Anforderungen des SFB 1313 geht jedoch hervor, dass Dateien von bis zu 150 GB anfallen können (siehe [Tabelle 1](#)). Dieser Herausforderung wird mit verschiedenen Strategien begegnet: Am Anfang steht die Frage, ob es überhaupt notwendig ist, die Daten in das zentrale Repositorium hochzuladen.

Sind die Daten leicht zu reproduzieren, genügt eine Anleitung zur Reproduktion, die meist wesentlich weniger Speicherplatz benötigt. Für Daten, die eher selten geteilt und bereits dezentral sicher gespeichert sind, reicht ein Link/Beschreibung des Speicherorts in den Metadaten.

Ist es notwendig, die Daten in DaRUS hochzuladen, sollte die Weboberfläche nur bis zu einer Dateigröße von 2 GB verwendet werden. Für größere Dateien bis zu 100 GB kann die Dataverse eigene REST-API mit entsprechend längeren Timeout-Einstellungen benutzt werden. Für Dateien > 100 GB ist derzeit eine dritte Strategie geplant, die das Hochladen *vorbei* am Webserver realisiert. Dafür sind Einmal-URLs vorgesehen, die über die REST-API angefordert werden können. Nach Abschluss des Uploads wird dann die Registrierung der Datei in DaRUS getriggert.

⁴<https://www.dataverse.org>. Letzter Zugriff 30.04.2019

4.2 Beschreibung der Daten - EngMeta

Für die Auffindbarkeit und Nachvollziehbarkeit der oben beschriebenen Daten müssen diese strukturiert so beschrieben werden, dass aus der Beschreibung die verwendeten Verfahren mit ihren Parametern, eingesetzte Hard- und Software und auch die betrachteten Variablen hervorgehen. EngMeta⁵ ist ein Schema, das neben allgemeinen deskriptiven Metadaten und technischen Informationen auch Beschreibungskategorien zum betrachteten System und zum Entstehungsprozess der Daten bietet. EngMeta beruht auf bestehenden Metadatenschemata wie DataCite, ExptML, CodeMeta und PROV und ergänzt diese um spezifische Felder.⁶ Es wurde an der Universität Stuttgart im Rahmen des BMBF geförderten Projektes DIPL-ING⁷ entwickelt und dient der umfassenden Beschreibung von Forschungsdaten aus den Ingenieurwissenschaften.

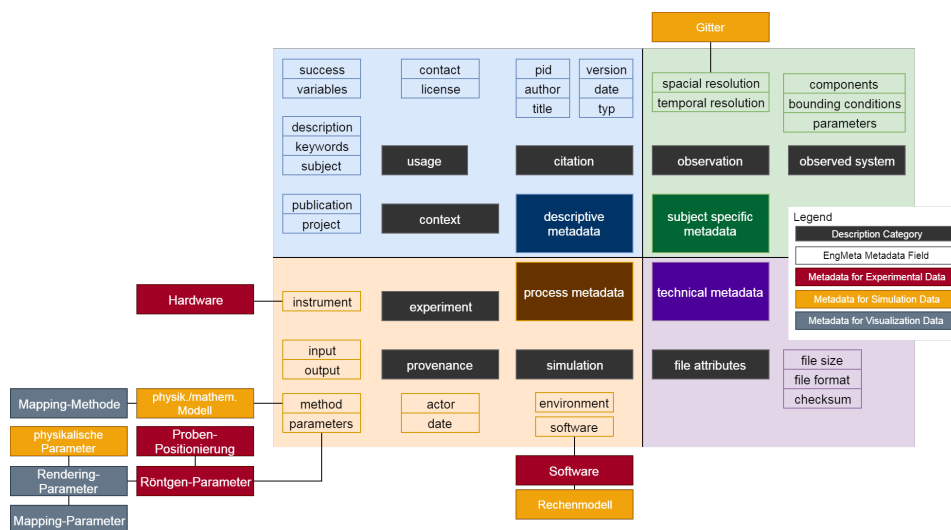


Abbildung 2: EngMeta

Abbildung 2 zeigt beispielhaft das Mapping der oben genannten Beschreibungskategorien von Experimental-, Simulations- und Visualisierungsdaten auf die entsprechenden Felder von EngMeta. Weitere wichtige Felder von EngMeta ermöglichen die Beschreibung der erhobenen oder kontrollierten Variablen (variables), die Komponenten der Porösen-Medien-Systeme (components) und die räumliche und zeitliche Auflösung der Datenerhebung (spacial/temporal resolution).

⁵<https://www.izus.uni-stuttgart.de/fokus/engmeta/>. Letzter Zugriff 01.07.2019

⁶Björn Schembera und Dorothea Iglezakis. 2019. "The Genesis of EngMeta - A Metadata Model for Research Data in Computational Engineering." Leading in a VUCA World (2019): S. 127–132. https://doi.org/10.1007/978-3-030-14401-2_12

⁷https://www.ub.uni-stuttgart.de/forschen-publizieren/forschungsdatenmanagement/projekte/dipl_ing/index.html. Letzter Zugriff 01.07.2019

4.3 Automatisierte Metadatenerfassung

Durch die Einführung einer so umfassenden Datenbeschreibung wie EngMeta ist eine manuelle Annotation der Daten nur mit hohem Zeitaufwand möglich. Um den Aufwand für die Beschreibung durch Metadaten und damit die Hemmschwelle zur Metadatenannotation zu senken, wird im Rahmen des Projektes DIPL-ING⁸ eine automatisierte Metadatenerfassung entwickelt. Durch diese Entwicklung können insbesondere die technischen und die fachspezifischen Metadaten sowie die Prozessmetadaten automatisiert erfasst werden: Technische Metadaten wie Dateigröße und -format stehen über die Dateiattribute bereit und lassen sich einfach automatisiert erfassen. Prozessmetadaten und fachspezifische Metadaten sind oft über die Ausgabe- oder Logdateien der Simulations- und Visualisierungs-codes wie auch der experimentellen Geräte verfügbar. Deskriptive Metadaten können über einheitliche Template-Dateien zumindest teilautomatisiert erfasst werden. Bezogen auf den Anwendungsfall bedeutet dies für die Simulationsdaten insbesondere die Erfassung von fachspezifischen Metadaten wie Simulationsparametern. Außerdem ist die Erfassung der physikalischen Parameter sowie weiterer Eigenschaften, wie z.B. dem Diskretisierungsverfahren, geplant. Darüber hinaus lassen sich auch Prozessmetadaten, wie Informationen über die genutzte Software oder die Rechnerumgebung erfassen. Bezogen auf die Experimentaldaten bedeutet dies die Erfassung von Prozess- und fachspezifischen Metadaten, die mit den Experimenten mitgeneriert werden.

Die Erfassung erfolgt schließlich mittels eines Tools, das sich über eine Konfigurationsdatei auf das jeweilige Ausgabeformat des Simulationscodes anpassen lässt. Ziel ist also die Entwicklung einer passenden Konfigurationsdatei zur Erfassung der oben genannten Metainformationen. Die automatisierte Metadatenerkennung durchsucht schließlich ein Verzeichnis und liest aus allen in der Konfigurationsdatei angegebenen Datei(-arten) die gewünschten Informationen aus und überträgt sie in das EngMeta-Format. In einem zweiten Schritt können die so bereits strukturierten Metadaten automatisiert in DaRUS eingelesen werden.

4.4 Visualisierung

Visualisierung ist ein grundlegender Bestandteil der Datenanalyse. In dem Anwendungskontext des Datenrepositoriums soll sie insbesondere eine schnelle Voransicht der Daten erlauben, ohne dass (potenzielle große) Datenmengen hierfür erst heruntergeladen und vom Benutzer selbst entsprechend verarbeitet werden müssen. Konzeptionell werden hierbei neue Visualisierungsdaten (z.B. Bilder und Videos) aus Rohdaten erzeugt, ihnen entsprechend zugeordnet und bereitgestellt.

Hierfür wird ein Ansatz für die automatische Erzeugung eines Ensembles von Visualisierungen für die a posteriori Analyse entwickelt. Es werden zunächst verschiedene

⁸https://www.ub.uni-stuttgart.de/forschen-publizieren/forschungsdatenmanagement/projekte/dipl_ing/index.html. Letzter Zugriff 01.07.2019

Visualisierungstechniken für die Abbildung der Rohdaten auf eine visuelle Repräsentation ausgewählt. Dies erfolgt automatisch auf Basis der entsprechenden Metadaten. Hierbei werden insbesondere die Datendomäne (2D, 3D, Zeitabhängigkeit, etc.) sowie die Datentypen (z.B. skalar bei Druck, vektoriell bei Geschwindigkeiten, etc.) berücksichtigt. Beispielsweise kommen für skalare Daten in einer 3D Domäne die Extraktion und das Rendering von Isoflächen in Betracht, ebenso wie das direkte Anzeigen der Volumen durch Raycasting. Für vektorielle Daten können integrale Linien berechnet werden, oder auch Glyphen (im einfachsten Fall Pfeile) verwendet werden.

Die ausgewählten Visualisierungstechniken weisen typischerweise zudem eine starke Parameterabhängigkeit auf. Dies betrifft sowohl die visuelle Repräsentation (z.B. Isowerte für Isoflächen, Transferfunktion für das Volumenrendering, Seedpunkte für die Erzeugung integraler Linien), als auch die Renderingparameter mit denen Bilder aus der Repräsentation erzeugt werden (insbesondere Kameraeinstellungen). Daraus ergibt sich folgender Visualisierungsparameterraum: Mappingmethode (kategorisch) \times Mappingparameter (kontinuierlich) \times Renderingparameter (kontinuierlich). Das Ergebnis einer konkreten Kombination bzw. eines Parametertupels ist ein Bild für einen Zeitschritt bzw. ein Video für zeitabhängige Daten. Eine angemessene Abtastung des Raums ergibt folglich eine Ansammlung an Bildern bzw. Videos, die für eine (eingeschränkte) Datenexploration zur Verfügung stehen. Ein wichtiger Vorteil dieser Lösung ist, dass für die Datenansicht kein direkter Zugriff auf die Originaldaten vonnöten ist.

5 Fazit und Ausblick

In diesem Artikel wurde ein Überblick über das Forschungsdatenmanagement des SFB 1313 gegeben. Hierbei stellt DaRUS die zentrale Plattform zur Verwaltung und Beschreibung von Daten dar. Zur Beschreibung der Metadaten wurde EngMeta entwickelt. EngMeta beruht auf bestehenden Metadatenschemata wie DataCite, ExptML, CodeMeta und PROV und ergänzt diese um spezifische Felder, die zur Datenerfassung im SFB benötigt werden. In diesem Rahmen wurden auch ein Tool zur automatisierten Metadatenerfassung sowie erste Visualisierungsansätze entwickelt.

Bei dem oben beschriebenen Visualisierungsansatz erfolgen die Erzeugung der Visualisierung und deren Analyse durch eine Anwenderin bzw. einen Anwender strikt getrennt voneinander. Die Exploration der Daten ist somit nur eingeschränkt möglich. Um dies flexibler zu gestalten, könnte zusätzlich ein Plugin entwickelt werden, das die interaktive Remote-Visualisierung im Browser erlaubt, mit der Möglichkeit z.B. die Kameraposition anzupassen. Dies erfordert konzeptionell jedoch den direkten Zugriff auf die Rohdaten für die freie Exploration. Dieser Ansatz stellt somit deutlich höhere Anforderungen sowohl an die technische Infrastruktur als auch die (effiziente) Umsetzung, um Interaktivität auch für mehrere Nutzer gleichzeitig gewährleisten zu können.

Zusätzlich muss eine Lösung für sehr große Datenmengen gefunden werden. Als ersten Schritt soll dies durch Verlinkung auf ein lokales Speicherbackend gelöst werden, wobei die Daten zunächst nicht bewegt werden. Bei publizierten Datensätzen muss jedoch die Verfügbarkeit gewährleistet werden, weshalb die Daten spätestens zur Publikation unverfälschbar auf einem sicheren Speichermedium mit direktem Zugriff von DaRUS abgelegt werden müssen. Die automatisierte Metadatenerfassung ist momentan in der Testphase und wird für verschiedene Softwarelösungen getestet.

Im Hinblick auf "Data-Integrated Simulation Sciences" soll DaRUS als Schnittstelle zwischen Simulation und Experiment dienen. Hierfür muss der effiziente Zugriff, sowie die Filterung von Metadaten gewährleistet werden.

Die genannten Punkte sind aktuelle Forschungsschwerpunkte des Datenmanagements im SFB 1313.